Ανάλυση Υπολοίπων

Τα υπόλοιπα για το πολλαπλό γραμμικό μοντέλο της παλινδρόμησης ορίζονται από τη σχέση $\underline{e} = \underline{\mathrm{Y}} - \underline{\hat{\mathrm{Y}}} = \underline{\mathrm{Y}} - \mathbf{X}\cdot\underline{\hat{\beta}} = (\mathbf{I}-\mathbf{P})\cdot\underline{\mathrm{Y}}$, όπου $\mathbf{P} = \mathbf{X}\cdot\left(\mathbf{X}^{\mathrm{T}}\cdot\mathbf{X}\right)^{-1}\cdot\mathbf{X}^{\mathrm{T}}$.

Ιδιότητες των υπολοίπων:

(1) $\mathrm{E}(\underline{e}) = \underline{0}$

(απόδειξη)

$$\mathrm{E}(\underline{e}) = \mathrm{E}\left[(\mathbf{I}-\mathbf{P})\cdot\underline{\mathrm{Y}}\right] = (\mathbf{I}-\mathbf{P})\cdot\mathrm{E}(\underline{\mathrm{Y}}) = (\mathbf{I}-\mathbf{P})\cdot\mathbf{X}\cdot\underline{\beta} = (\mathbf{I}\cdot\mathbf{X}-\mathbf{P}\cdot\mathbf{X})\cdot\underline{\beta} = \mathbf{0}\cdot\underline{\beta}$$

(2) $Cov(\underline{e}) = \sigma^2\cdot(\mathbf{I}-\mathbf{P})$

(απόδειξη)

$$Cov(\underline{e}) = Cov\left[(\mathbf{I}-\mathbf{P})\cdot\underline{\mathrm{Y}}\right] = (\mathbf{I}-\mathbf{P})\cdot Cov(\underline{\mathrm{Y}})\cdot(\mathbf{I}-\mathbf{P})^{\mathrm{T}} = (\mathbf{I}-\mathbf{P})\cdot\sigma^2\cdot\mathbf{I}\cdot(\mathbf{I}-\mathbf{P})^{\mathrm{T}} =$$
$$= \sigma^2\cdot(\mathbf{I}-\mathbf{P})\cdot(\mathbf{I}-\mathbf{P})^{\mathrm{T}} = \sigma^2\cdot(\mathbf{I}-\mathbf{P})\cdot(\mathbf{I}-\mathbf{P}) = \sigma^2\cdot(\mathbf{I}-\mathbf{P})^2 = \sigma^2\cdot(\mathbf{I}-\mathbf{P})$$

Άρα συμπεραίνουμε ότι $Var(e_i) = \sigma^2\cdot(1-\mathrm{P}_{ii})$ και $Cov(e_i, e_j) = -\sigma^2\cdot\mathrm{P}_{ij}$, $i \neq j$, όπου $\mathbf{P} = (\mathrm{P}_{ij})$, $i = 1,2,\ldots,n$, $j = 1,2,\ldots,n$.

(3) Κατανομή των υπολοίπων:

$\underline{e} \sim \mathrm{N}_n\left(\underline{0}, \sigma^2\cdot(\mathbf{I}-\mathbf{P})\right)$

(απόδειξη)

Αφού $\underline{e} = (\mathbf{I}-\mathbf{P})\cdot\underline{\mathrm{Y}}$ είναι γραμμικός συνδυασμός κανονικών.

(4) $Cov(\underline{e}, \underline{\hat{\mathrm{Y}}}) = 0$

(απόδειξη)

$$Cov(\underline{e}, \underline{\hat{\mathrm{Y}}}) = Cov\left((\mathbf{I}-\mathbf{P})\cdot\underline{\mathrm{Y}}, \mathbf{P}\cdot\underline{\mathrm{Y}}\right) = (\mathbf{I}-\mathbf{P})\cdot Cov(\underline{\mathrm{Y}}, \underline{\mathrm{Y}})\cdot\mathbf{P} = (\mathbf{I}-\mathbf{P})\cdot Cov(\underline{\mathrm{Y}})\cdot\mathbf{P} =$$
$$= (\mathbf{I}-\mathbf{P})\cdot\sigma^2\cdot\mathbf{I}\cdot\mathbf{P} = \sigma^2\cdot(\mathbf{I}-\mathbf{P})\cdot\mathbf{P} = \sigma^2\cdot(\mathbf{P}-\mathbf{P}^2) = \sigma^2\cdot(\mathbf{P}-\mathbf{P}) = \sigma^2\cdot\mathbf{0} = \mathbf{0}$$

Άρα τα υπόλοιπα και οι προβλέψεις είναι ασυσχέτιστα και όταν ισχύει η κανονικότητα, είναι και ανεξάρτητες τυχαίες μεταβλητές.

(5) $Cov(\underline{e}, \underline{\mathrm{Y}}) = \sigma^2\cdot(\mathbf{I}-\mathbf{P})$

(απόδειξη)

$$Cov(\underline{e}, \underline{\mathrm{Y}}) = Cov\left((\mathbf{I}-\mathbf{P})\cdot\underline{\mathrm{Y}}, \underline{\mathrm{Y}}\right) = (\mathbf{I}-\mathbf{P})\cdot Cov(\underline{\mathrm{Y}}, \underline{\mathrm{Y}}) = (\mathbf{I}-\mathbf{P})\cdot Cov(\underline{\mathrm{Y}}) = (\mathbf{I}-\mathbf{P})\cdot\sigma^2\cdot\mathbf{I} =$$
$$= \sigma^2\cdot(\mathbf{I}-\mathbf{P})$$

Δύο άλλες μορφές υπολοίπων:

(α) Τυποποιημένα Υπόλοιπα (standardized residuals):

$$e_{si} = \frac{e_i}{\hat{\sigma}} = \frac{\mathrm{Y}_i - \hat{\mathrm{Y}}_i}{\hat{\sigma}}$$

Τα τυποποιημένα υπόλοιπα δεν εξαρτόνται από τη μονάδα μέτρησης και δείχνουν το μέγεθος των e_i σε σχέση με $\hat{\sigma}$.

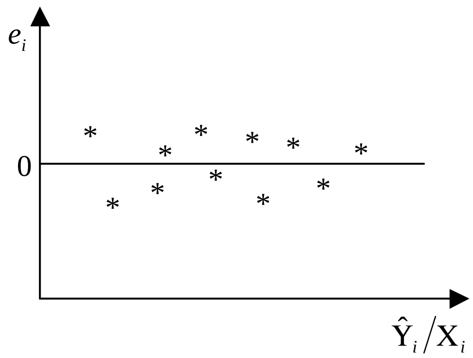
(β) Μαθηματικοποιημένα Υπόλοιπα (studentized residuals):

$$e_{sti} = \frac{e_i}{\hat{\sigma} \cdot \sqrt{1 - \mathrm{P}_{ii}}}$$

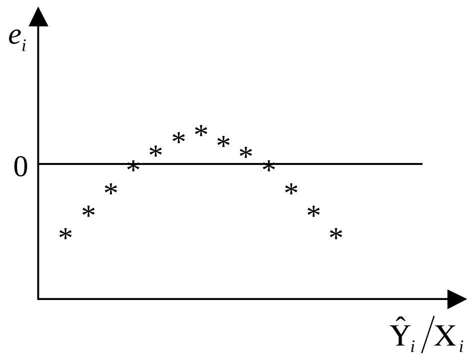
Ανάλυση Υπολοίπων

Η ανάλυση υπολοίπων είναι μια σειρά μεθόδων (γραφικών και στατιστικών) που μας επιτρέπει να ελέγξουμε τις υποθέσεις για τα σφάλματα, καθώς και την ορθότητα ή μη του μοντέλου.

Α) Έλεγχος Ύπαρξης Ασυσχέτιστων Σφαλμάτων:
Μια γραφική μέθοδος είναι η εξής:
Κατασκευάζουμε γραφική παράσταση των υπολοίπων e_i ως προς $\hat{Y}_i$ ή/και X_i, επειδή τα σφάλματα είναι ασυσχέτιστα με $\hat{Y}_i$ και X_i. Δεν κατασκευάζω γραφική παράσταση των e_i ως προς Y_i, γιατί δεν είναι ασυσχέτιστα.



(1)

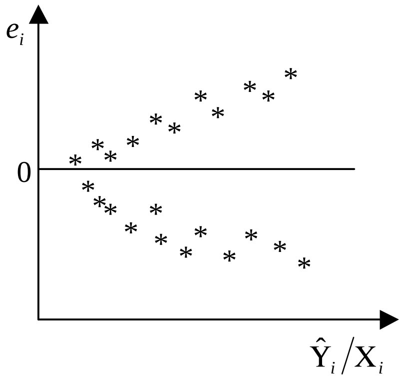


(2)

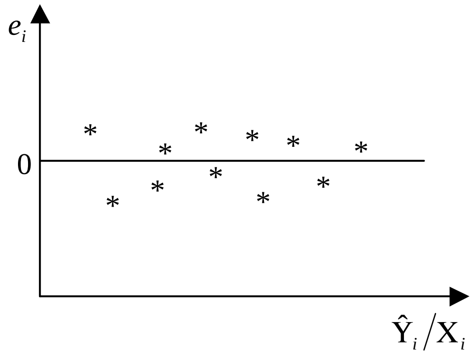
Στο (2) έχω ένδειξη συσχτισμένων σφαλμάτων, ενώ στο (1) έχω ένδειξη ασυσχέτιστων σφαλμάτων.

Ένας στατιστικός έλεγχος είναι ο παρακάτω: (έλεγχος των ροών)
H_0 : η ακολουθία των “+” στα υπόλοια είναι τυχαία (ασυσχέτιστα σφάλματα)
H_1 : η ακολουθία των “+” στα υπόλοια δεν είναι τυχαία

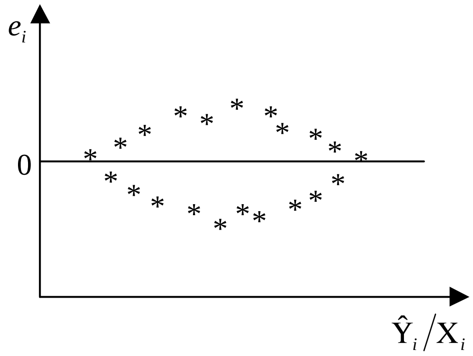
Β) Έλεγχος Σταθερής Διακύμανσης: ($Var(\varepsilon_i) = \sigma^2, i = 1,2,...,n$)
Σταθερή διακύμανση συνεπάγεται ότι η διακύμανση των σφαλμάτων δεν εξαρτάται από τα X_i (τιμές της ανεξάρτητης μεταβλητής).



Για μεγάλες τιμές των X_i έχω μεγάλες αποκλίσεις των σφαλμάτων. Ένδειξη μη σταθερής διακύμανσης.



Εδώ βλέπουμε σταθερή διακύμανση.



Για μεσαίες τιμές των X_i έχω μεγάλες αποκλίσεις. Για ακραίες τιμές έχω μικρές αποκλίσεις. Ένδειξη μη σταθερής διακύμανσης.

<u>Διόρθωση μη σταθερής διακύμανσης:</u>
Υπάρχουν οι εξής δύο μεθόδοι:
(1) Μετασχηματισμός σταθεροποίησης της διακύμανσης.
(2) Γενικευμένοι εκτιμητές ελαχίστων τετραγώνων.

<u>Μέθοδος #1:</u>
Έστω το απλό γραμμικό μοντέλο $Y_i = \beta_0 + \beta_1 \cdot X_i + \varepsilon_i$, $i = 1,2,...,n$, με $E(\varepsilon_i) = 0$, $Cov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$, και $Var(\varepsilon_i) = \sigma^2 \cdot X_i$ (εξαρτάται από X_i).

Χρησιμοποιούμε τις εκτιμήτριες ελαχίστων τετραγώνων $\hat{\beta}_0 = \overline{\mathrm{Y}} - \hat{\beta}_1 \cdot \overline{\mathrm{X}}$ και $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} (\mathrm{X}_i - \overline{\mathrm{X}}) \cdot (\mathrm{Y}_i - \overline{\mathrm{Y}})}{\sum_{i=1}^{n} (\mathrm{X}_i - \overline{\mathrm{X}})^2}$, αλλά οι ιδιότητες τους δεν ισχύουν πλεόν ($\hat{\beta}_0, \hat{\beta}_1$ δεν είναι ΑΟΕΔ όταν ισχύει η κανονικότητα).

Εφαρμόζω κάποιο μετασχηματισμό για να βρώ άλλο εκτιμητή:

Αν όλα τα $\mathrm{X}_i \neq 0$, $\dfrac{\mathrm{Y}_i}{\mathrm{X}_i} = \dfrac{\beta_0}{\mathrm{X}_i} + \beta_1 + \dfrac{\varepsilon_i}{\mathrm{X}_i} \Rightarrow \mathrm{Y}_i^* = \beta_0^* + \beta_1^* \cdot \mathrm{X}_i^* + \varepsilon_i^*$,

όπου $\mathrm{Y}_i^* = \dfrac{\mathrm{Y}_i}{\mathrm{X}_i}$, $\mathrm{X}_i^* = \dfrac{1}{\mathrm{X}_i}$, $\beta_0^* = \beta_1$, $\beta_1^* = \beta_0$, $\varepsilon_i^* = \dfrac{\varepsilon_i}{\mathrm{X}_i}$.

Αλλά τώρα έχω το μοντέλο $\mathrm{Y}_i^* = \beta_0^* + \beta_1^* \cdot \mathrm{X}_i^* + \varepsilon_i^*$, με

$$\left. \begin{array}{l} \mathrm{E}(\varepsilon_i^*) = 0 \\ Var(\varepsilon_i^*) = \sigma^2 \\ Cov(\varepsilon_i^*, \varepsilon_j^*) = 0, \ i \neq j \\ \varepsilon_i^* \sim \mathrm{N}(0, \sigma^2) \end{array} \right\} \quad \text{κλασσικές υποθέσεις του γραμμικού μοντέλου}$$

Επομένως, οι εκτιμήτριες ελαχίστων τετραγώνων $\hat{\beta}_0^*, \hat{\beta}_1^*$ έχουν όλες τις βέλτιστες ιδιότητες και συνεπώς $\hat{\beta}_1 = \hat{\beta}_0^*$ και $\hat{\beta}_0 = \hat{\beta}_1^*$.

<u>Μέθοδος #2:</u>

Γενικά μπορούμε να θεωρήσουμε το πολλαπλό μοντέλο παλινδρόμησης $\underline{\mathrm{Y}} = \mathbf{X} \cdot \underline{\beta} + \underline{\varepsilon}$, $\mathrm{E}(\underline{\varepsilon}) = \underline{0}$, $Cov(\underline{\varepsilon}) = \sigma^2 \cdot \Sigma$, $\Sigma > 0$, όπου Σ γνωστός $(n \times n)$ πίνακας. Άρα τώρα έχω συσχετισμένα σφάλματα.

Αφού $\Sigma > 0 \Rightarrow \exists \Sigma^{-1} > 0 \Rightarrow \exists \Sigma^{-1/2} : \Sigma^{-1/2} \cdot \Sigma^{-1/2} = \Sigma^{-1}$.

Οπότε $\Sigma^{-1/2} \cdot \underline{\mathrm{Y}} = \Sigma^{-1/2} \cdot \mathbf{X} \cdot \underline{\beta} + \Sigma^{-1/2} \cdot \underline{\varepsilon} \Rightarrow \underline{\mathrm{Y}}^* = \mathbf{X}^* \cdot \underline{\beta} + \underline{\varepsilon}^*$,

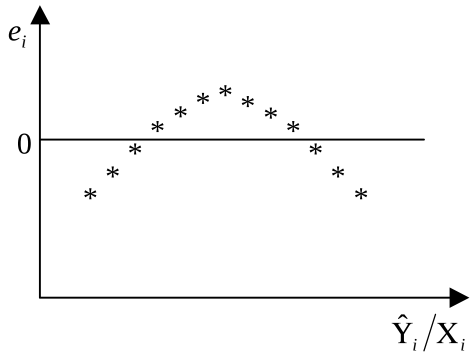
όπου $\underline{\varepsilon}^* \sim \mathrm{N}_n(\underline{0}, \sigma^2 \cdot \mathbf{I})$, όταν ισχύει η κανονικότητα.

Συνεπώς, $\hat{\underline{\beta}} = (\mathbf{X}^{*\mathrm{T}} \cdot \mathbf{X}^*)^{-1} \cdot \mathbf{X}^{*\mathrm{T}} \cdot \underline{\mathrm{Y}}^* = (\mathbf{X}^{\mathrm{T}} \cdot \Sigma^{-1/2} \cdot \Sigma^{-1/2} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \Sigma^{-1/2} \cdot \Sigma^{-1/2} \cdot \underline{\mathrm{Y}} =$
$= (\mathbf{X}^{\mathrm{T}} \cdot \Sigma^{-1} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \Sigma^{-1} \cdot \underline{\mathrm{Y}}$ (γενικευμένη εκτιμήτρια ελαχίστων τετραγώνων)

Η γενικευμένη εκτιμήτρια ελαχίστων τετραγώνων έχει όλες τις βέλτιστες ιδιότητες, π.χ. $\hat{\underline{\beta}}$ είναι ΑΟΕΔ.

Γ) Έλεγχος Ορθότητας Γραμμικού Μοντέλου:

Γραμμική μέθοδος: Γραφική παράσταση των υπολοίπων ως προς X_i ή ως προς $\hat{Y}_i$.



Παρατηρείται ο σχηματισμός κάποιας καμπύλης. Πιθανώς να πρέπει να εισαχθεί κάποιος δευτεροβάθμιος όρος ή τα σφάλματα να είναι συσχετισμένα.

Δηλαδή: Ενώ υποθέτουμε ότι ισχύει το γραμμικό μοντέλο $Y = \beta_0 + \beta_1 \cdot X + \varepsilon$, το πιο πάνω σχήμα δίνει ένδειξη του μοντέλου $Y = \beta_0 + \beta_1 \cdot X + \beta_2 \cdot X^2 + \varepsilon$.

Εξήγηση: Αφού $Y - \hat{Y} = \left(\beta_0 - \hat{\beta}_0\right) + \left(\beta_1 - \hat{\beta}_1\right) \cdot X$, δεν θα πρέπει να έχω καμία συστηματική εμφάνιση στο σχήμα. Αλλά αν ισχύει το δευτεροβάθμιο μοντέλο, $Y - \hat{Y} = \left(\beta_0 - \hat{\beta}_0\right) + \left(\beta_1 - \hat{\beta}_1\right) \cdot X + \beta_2 \cdot X^2$.

Παρατήρηση:
Συχνά, ένα μη γραμμικό μοντέλο μοντέλο μπορεί να μετασχηματιστεί σε γραμμικό.

π.χ. $Y = \alpha \cdot e^{\beta \cdot X} \Rightarrow \log(Y) = \log(\alpha) + \beta \cdot X$

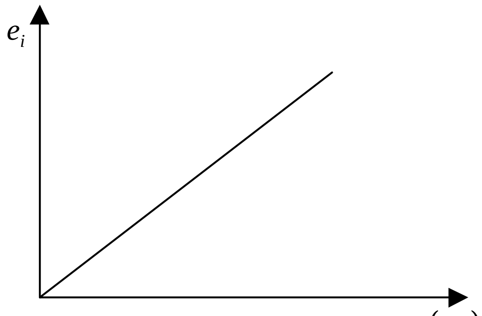
$Y = \alpha \cdot X^{\rho} \Rightarrow \log(Y) = \log(\alpha) + \rho \cdot \log(X)$

Δ) Έλεγχος Κανονικότητας Σφαλμάτων:
Αυτός ο έλεγχος εκτελείται με έλεγχο κανονικότητας των υπολοίπων ή με γραφική παράσταση (Q-Q plot).

Αν η τ.μ. X είναι κανονική, τότε $P\left(X \le x_p\right) = 1 - p \Rightarrow P\left(Z \le \dfrac{x_p - \mu}{\sigma}\right) = 1 - p$.

Άρα $\dfrac{x_p - \mu}{\sigma} = z_p \Rightarrow x_p = \mu + \sigma \cdot z_p$.

Συνεπώς,



Θα έχω ευθεία γραμμή αν το δείγμα προέρχεται από την κανονική κατανομή.

<u>Παρατήρηση:</u>
Αν κάνουμε γραφικές παραστάσεις των e_{si} ή e_{sti}, καταλήγουμε στα ίδια συμπεράσματα.

<u>Ακραίες Παρατηρήσεις</u>

Η παρατήρηση Y_j (j συγκεκριμένο) είναι ακραία αν το αντίστοιχο υπόλοιπο $e_j = \mathrm{Y}_j - \hat{\mathrm{Y}}_j$ είναι κατά πολύ μεγαλύτερο (κατά απόλυτη τιμή) από όλα τα άλλα υπόλοιπα.

<u>Μεθοδολογία για την ανεύρεση ακραίας παρατήρησης:</u>

Α) Παραλείπουμε από την ανάλυση την j παρατήρηση. Έτσι έχουμε $(n-1)$ παρατηρήσεις.

Β) Σχηματίζουμε τις εκτιμήτριες $\underline{\hat{\beta}}_{(-j)}, \hat{\sigma}^2_{(-j)}$, που στηρίζονται στις $(n-1)$ παρατηρήσεις.

Γ) Σχηματίζουμε τη διαφορά $\mathrm{Y}_j - \hat{\mathrm{Y}}_{(-j)}$, όπου η $\hat{\mathrm{Y}}_{(-j)}$ προκύπτει από τις $(n-1)$ παρατηρήσεις, δηλαδή $\hat{\mathrm{Y}}_{(-j)} = \underline{\mathrm{X}}_j^{\mathrm{T}} \cdot \underline{\hat{\beta}}_{(-j)}$, όπου $\underline{\mathrm{X}}_j^{\mathrm{T}}$ η j γραμμή του πίνακα $\mathbf{X}$.

Δ) Ελέγχω αν η j παρατήρηση είναι ακραία με ελεγχοσυνάρτηση $t = \dfrac{\mathrm{Y}_i - \hat{\mathrm{Y}}_{(-i)}}{\hat{\sigma}_{(-i)} \cdot \sqrt{\mathrm{A}}} \sim t_{(n-1)-(p+1)}$, όπου $\mathrm{A} = \left[1 + \underline{\mathrm{X}}_j^{\mathrm{T}} \cdot \left(\mathbf{X}^{\mathrm{T}}_{(-j)} \cdot \mathbf{X}_{(-j)}\right)^{-1} \cdot \underline{\mathrm{X}}_j\right]$.

Απορρίπτουμε την υπόθεση H_0 : η j παρατήρηση δεν είναι ακραία, σε επίπεδο σημαντικότητας α , αν $|t| > t_{(n-1)-(p+1),\alpha/2}$.